



Supporting Collaborative Modeling via Natural Language Processing

Fatma Başak Aydemir¹ Fabiano Dalpiaz²

¹Boğaziçi University 

²Utrecht University 

39th International Conference on Conceptual Modeling



Utrecht University

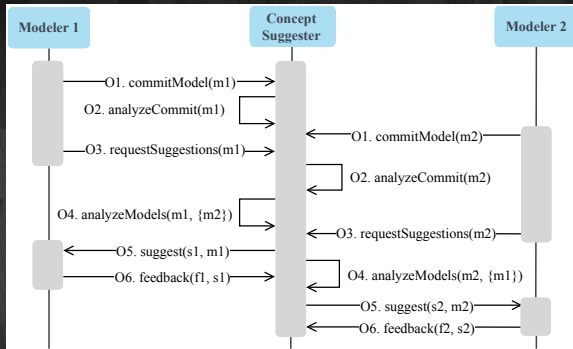
Collaborative Modeling

Various

- Experts
- Modeling Languages
- Locations
- Time Zones



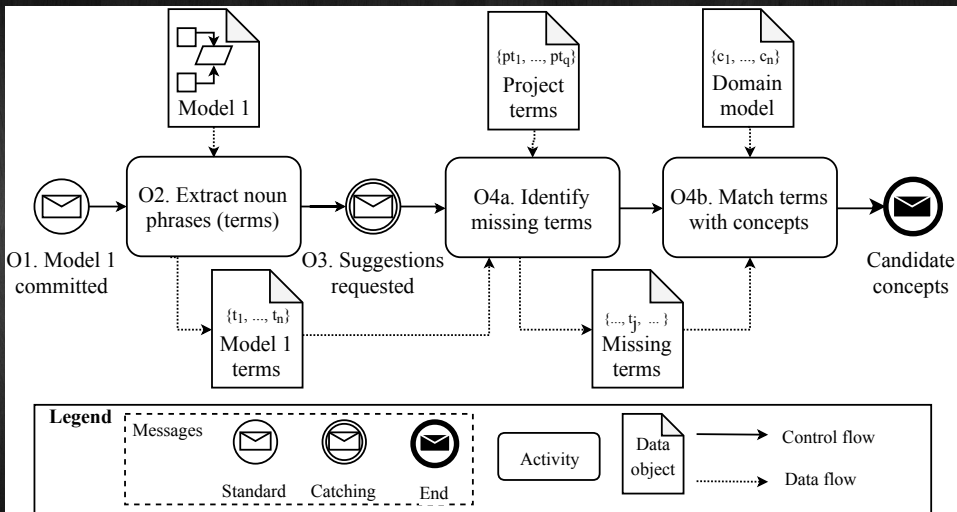
Concept Suggestion Service



A web service integrated to the modeling environments

- Ignores the meta-models
- Analyzes the labels
- Keeps track of the modelled concepts
- Suggests missing concepts to modelers
- Hides model details
- Completeness
- Common vocabulary

Service Operations



Matching Heuristics

Possible matches for “thesis”

- Exact match
- Sub-string match
- Similarity
- Relatedness

Matching Heuristics

- Exact match
- Sub-string match
- Similarity
- Relatedness

Possible matches for “thesis”

- Thesis

Matching Heuristics

- Exact match
- Sub-string match
- Similarity
- Relatedness

Possible matches for “thesis”

- Thesis
- Thesis Project

Matching Heuristics

- Exact match
- Sub-string match
- Similarity
- Relatedness

Possible matches for “thesis”

- Thesis
- Thesis Project
- Graduation project

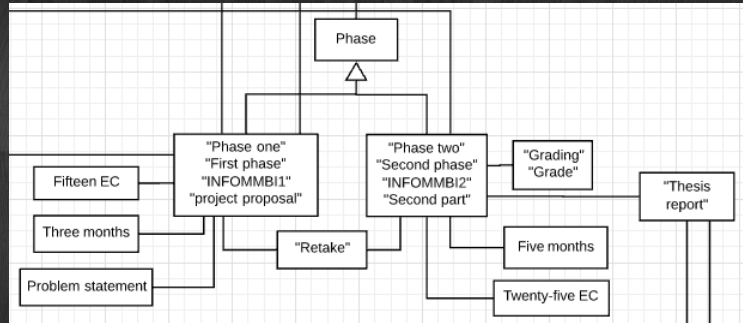
Matching Heuristics

- Exact match
- Sub-string match
- Similarity
- Relatedness

Possible matches for “thesis”

- Thesis
- Thesis Project
- Graduation project
- Coordinator

Suggestion Heuristics



Filtering Heuristics

- Fixed number
- User feedback
- Frequency
- Limiting matches per missing item

Similarity of Compound Nouns

- Compound nouns for many concepts
- Used in heuristics
 - Add detail to the models
 - Similarity check for common terminology
 - Explore the domain model
- Two algorithms to calculate the similarity of a pair of compound nouns
 - WordNet and Word2Vecbased

Similarity of Compound Nouns

- Compound nouns for many concepts
- Used in heuristics
 - Add detail to the models
 - Similarity check for common terminology
 - Explore the domain model
- Two algorithms to calculate the similarity of a pair of compound nouns
 - WordNet and Word2Vecbased
 - Domain model based

WordNet and Word2Vec based Similarity

- Get the Word2Vec similarity scores of noun pairs of compound words
- If the score is higher than a threshold check if they are synonyms using WordNet
- Set the score to 1 for synonyms, leave as is otherwise
- The similarity of the compounds is a weighted average of the pairs

Thesis Project, Graduation Project

$$\gamma \cdot \text{sim}(\text{thesis}, \text{graduation}) + \delta \cdot \text{sim}(\text{project}, \text{project}) + \epsilon \cdot \text{sim}(\text{thesis}, \text{project}) + \kappa \cdot \text{sim}(\text{project}, \text{graduation}) \quad (1)$$

Domain Model based Similarity

- Based on how well the compounds are matched in the domain model
- Individual matching score for each compound is calculated
- The similarity of the two term is the average of their scores

Experimental Setup to Detect Similarity

- Used 20 pairs of 2-word compound nouns
- Gold standard: surveyed people to assess the similarity on a 5-point Likert type scale
- Compared the results with Bert-web and spaCy similarity

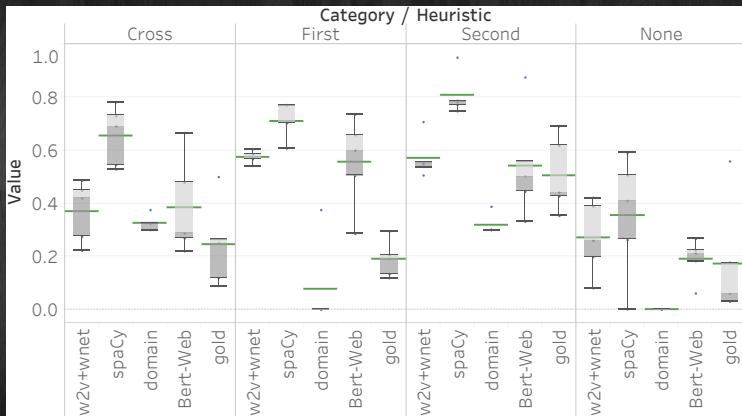
Pairs Used

Pair ID	First compound	Second compound	Cat.	Pair ID	First compound	Second compound	Cat.
P1	thesis project	project facilitator	1	P11	company supervisor	second supervisor	3
P2	MBI student	student administration	1	P12	first supervisor	second supervisor	3
P3	MBI thesis	thesis topic	1	P13	computing science	information science	3
P4	graduation project	project idea	1	P14	project proposal	short proposal	3
P5	literature review	relevant literature	1	P15	official ceremony	graduation ceremony	3
P6	first phase	first supervisor	2	P16	scientific paper	official ceremony	4
P7	second phase	second presentation	2	P17	Google calendar	MBI colloquium	4
P8	graduation ceremony	graduation supervisor	2	P18	company supervisor	project facilitator	4
P9	MBI thesis	MBI colloquium	2	P19	department member	participation token	4
P10	thesis topic	thesis report	2	P20	research question	literature review	4

Most and Least Similar Pairs

Word embeddings + WordNet	spaCy	Domain model	Gold standard
first supervisor - second supervisor	.71 first supervisor - second supervisor	.95 computing science - information science	.39 official ceremony - graduation ceremony
thesis topic - thesis report	.60 official ceremony - graduation ceremony	.79 MBI thesis - thesis topic	.38 project proposal - short proposal
second phase - second presentation	.59 project proposal - short proposal	.78 MBI thesis - MBI colloquium	.38 company supervisor - project facilitator
MBI thesis - thesis topic	.27 MBI thesis - thesis topic	.53 first phase - first supervisor	.00 thesis project - project facilitator
department member - participation token	.26 company supervisor - project facilitator	.51 second phase - second presentation	.00 first phase - first supervisor
MBI student - student administration	.22 department member - participation token	.41 graduation ceremony - graduation supervisor	.00 MBI student - student administration

Results i



Results ii

- spaCy has the highest Pearson correlation with the gold standard
- Domain model based approach has the least Euclidian distance to the gold standard
- Specific heuristics resembles the gold standard the most
- Bert has promising results as well

Discussion

- NLP can be used support collaborative modeling activities
- Conceptual modeling domain has its own challenges in terms of NLP
- We need more data and empirical validation for both the service and the similarity algorithms

Conclusions

- A lightweight NLP-powered service to facilitate collaborative modeling
- Heuristics to detect the concepts that are modelled, and to provide suggestions from a domain model for those concepts that are not modelled in a multi-model and multi-modeler collaborative modelling setting
- This work focuses on the challenge of identifying similar compound nouns
- Our domain model based algorithm performs well, but we need further validation to be conclusive

Contact us!



✉ basak.aydemir@boun.edu.tr
🐦 @aydemirfb



✉ f.dalpiaz@uu.nl
🐦 @FabianoDalpiaz

Thank you for your attention!